Appendix C

Empirical Relationship Between Use and Air Concentration

Empirical Relationship Between Use, Area, and Ambient Air Concentration of Methyl Bromide for Subchronic Exposure Concerns
Executive Summary

Background

This document "Empirical Relationship Between Use, Area, and Ambient Air Concentration of Methyl Bromide for Subchronic Exposure Concerns" is the final version of the air monitoring analysis distributed May 21, 2001. It incorporates corrections of errors identified by Department of Pesticide Regulation (DPR) staff and commenters, recalculated regression models using corrected data, identification of a new source of methyl bromide, and rewritten sections to clarify areas identified by commenters.

DPR analyzed methyl bromide air monitoring data conducted in 2000 in Kern, Monterey, and Santa Cruz counties and available methyl bromide soil use data in geographical areas in proximity to the study monitoring locations. The analysis utilized pesticide use report records collected on an accelerated basis in order to provide a timely evaluation. Other sources of information were not available under the time constraints necessary to evaluate the sources of the monitored air concentrations and prepare a mitigation strategy before the 2001 peak use season.

In order to determine the extent of contribution to monitored air concentrations of methyl bromide soil applications, DPR utilized a statistical regression analysis. This was used to measure the strength of the relationship (correlation, r), to determine the extent that the methyl bromide soil use data statistically explains the air monitoring data (determination, R^2), and to establish a statistical model that characterizes the relationship. The analysis characterizes methyl bromide use by section and evaluates proximity to each monitoring station by incrementally including additional sections in a symmetrical expansion from the one that includes the location of the monitoring site.

Results

Methyl bromide soil use in sections and the monitored air concentrations were significantly correlated, with few exceptions. Methyl bromide use characterized by section areas was most consistent in explaining the air monitoring data with all 24 of 24 models establishing a significant predictive relationship. This held true for monitoring periods of 1, 3-4 and 7-8 weeks when analyzed with the corresponding periods of methyl bromide soil use. The best predictive models were from the 7-8 week air monitoring period, inclusive of the whole study duration. These different area models indicate that methyl bromide soil use explains 67 to 95% of the variation in corresponding air concentrations, statistically speaking, with the best model being the 7 x 7 section area model. This model, explaining 95% of the variation in monitored air concentrations from methyl bromide soil uses in a 7 x 7 section area, leaves only an estimated 5% for contributing factors like weather, topography, or directional use patterns.

The best model, statistically speaking (7 x 7 area methyl bromide use for 7-8 weeks), provides a means to calculate resultant air concentrations from incremental methyl bromide soil use in an area slightly larger than a township. This may be useful when evaluating mitigation strategies. For example, setting the concentration equal to the reference level of 1 ppb, and solving for the corresponding use level, results in a use level of approximately 20,000 lbs per township per month. This use level was utilized to query the pesticide use report database for townships which exceeded the 20,000 lb for three consecutive months. This occurred in seventeen townships in 1999.

This report employed a straight forward statistical analysis to interpret its results. Further analyses using more sophisticated statistics should be beneficial and may explain some of the anomalies visible in portions of the analyses.

Empirical relationship between use, area, and ambient air concentration of methyl bromide for subchronic exposure concerns

LinYing Li, Bruce Johnson and Randy Segawa

Department of Pesticide Regulation
Environmental Monitoring Branch
June 2001

1. Introduction

The Department of Pesticide Regulation (DPR) recommended monitoring sites and periods for air sampling to the Air Resources Board (ARB) for monitoring ambient air concentrations of methyl bromide in Monterey/Santa Cruz and Kern Counties under the Toxic Air Contamination Program (AB1807) in year 2000 [1]. These recommendations specified areas of historically heavy use and times of peak use in these regions. The ARB conducted air sampling and lab analysis for six sites in Monterey/Santa Cruz counties and six sites in Kern County [2,3]. The results indicated that methyl bromide air concentrations in Monterey/Santa Cruz counties were generally higher than those in Kern County. The highest 24-hour concentration observed in Monterey/Santa Cruz was 28.28 ppb, well below the 1-day acute reference levels established by the Medical Toxicology Branch of DPR (210 ppb and 250 ppb for adults and children respectively [4]). However, the 8-week average concentrations at some sites exceeded the reference level of subchronic exposure (6 to 8 weeks time frame). The reference levels of human subchronic exposure are 2 ppb for adults and 1 ppb for children[4]. The highest 8 week average concentration was 7.73 ppb, exceeding the 1 ppb reference level.

Methyl bromide use pattern (application amount, frequency and density) near the monitoring site during the sampling period was perhaps the dominant factor that influenced air concentrations. This statistical analysis relates the measured air concentrations to the local methyl bromide use in various areas with the monitoring site as a centroid. The objectives are 1) to establish empirical relationships between air concentration and zonal use of methyl bromide; 2) to estimate the size of area surrounding a monitoring site where methyl bromide applications significantly affected the air concentrations; and 3) to provide the mechanism to estimate subchronic air concentrations as a function of use. This report documents the procedure and results of using statistical methods to analyze AB1807 methyl bromide data for year 2000.

2. Methods and Materials

2.1 Location of Monitoring Sites

Six sampling sites were selected in Monterey/Santa Cruz Counties, and six in Kern

County. For each monitoring site in Monterey/Santa Cruz Counties, ARB provided three references for its location: 1) GPS coordinates, 2) section/township/range (STR), and 3) street numbers of the institution (usually a school) where the monitoring station was sited. For Kern County, only 2) and 3) were provided. Among them the section/township/range of monitoring sites is of the most interest to this analysis, because pesticide applications are referenced by Meridian Township Range Section (MTRS) in the Pesticide Use Report (PUR). To make sure the MTRS is correct for each monitoring site, ArcView GIS was used to locate monitoring sites on the map. The three references did not always point to the same site on the map. Sometimes the difference was substantial. When inconsistencies occurred, other tools such as Yahoo! Map were also used to assist locating of sampling sites. In the situation that all these efforts could not resolve the difference, phone calls were made to people who work in the institution where the monitoring site was located to determine its location.

In ARB reports, the location of most monitoring sites in Monterey/Santa Cruz Counties was not accurate enough for this analysis. We checked the STR locations of all sites in Monterrey/Santa Cruz counties, 5 out of 6 sites were incorrectly annotated. Final determination was made in consultation with Monterey County Agricultural Commissioner (CAC) staff. Although no GPS coordinates were provided for monitoring sites in Kern County, we utilized street numbers and GIS to check the STR locations. Two sites were found incorrectly located in the ARB report. Kern County CAC staff assisted in verifying the STR location for Kern sites.

The original ARB locations and the DPR-determined locations, which were used for the calculations, are listed in Table 1. The final locations are mapped in Figures 1-3.

Table 1: Location of monitoring sites

County	Site	Section Township Range (STR)		
		DPR	ARB	Note
Kern	ARB	S.34/T.29S/R.27E	S.34/T.29S/R.27E	OK
Kern	CRS	S.33/T.27S/R.25E	S.33/T.27S/R.25E	OK
Kern	MET	S.1/T.11N./R.20W	S.1/T.11N./R.20W	OK
Kern	MVS	S.30/T.30S/R.29E	S.30/T.30S/R.28E	Changed
Kern	SHA	S.10/T.28S/R.25E	S.10/T.27S/R.25E	Changed
Kern	VSD	S.19/T.31S/R.29E	S.19/T.31S/R.29E	OK
Monterey	CHU	S.3/T.16S/R.4E	S.3/T.16S/R.3E	Changed
Monterey	LJE	S.10/T.14S/R.3E	S.10/T.13S/R.3E	Changed
Monterey	OAS	S.31/T.18S/R.7E	S.6/T.18S/R.6E	Changed
Monterey	SAL	S.22/T.14S/R.3E	S.27/T.14S/R.3E	Changed
Monterey	PMS	S.9/T.12S/R.2E	S.9/T.12S/R.1E	Changed
Santa Cruz	SES	S.22/T.11S/R.2E	S.22/T.11S/R.2E	Ok

2.2 Air Concentration

In Kern County, air sampling started on July 19th and ended on August 31st, lasting

7 weeks. Sampling in Monterey/Santa Cruz started from September 11th and ended on November 2nd, lasting about 8 weeks. The ARB provided daily average air concentration data in its summary reports [2,3] for this monitoring project. In general, the air sampling was conducted from Monday through Thursday. One exception was the first week in Kern County, with monitoring only on Wednesday and Thursday. The average air concentration over various time periods was calculated from the daily average air concentration data for each monitoring site. It was assumed that the 4-day average could represent the average air concentration for the week.

2.3 Methyl Bromide Use

Methyl bromide use surrounding the monitoring site was quantified in two ways: in an area, and in a tract (Fig.4). For example, in an area of 5X5 mile^2 centered on a section containing a monitoring site, methyl bromide use in pounds was summed over each week. Methyl bromide use amount was also calculated over a tract that is, say, roughly 3 miles away (tract 3) from the monitoring site. Tract 0 consisted of the single section containing the monitoring site. Tract 1 consisted of the eight sections surrounding, but not including tract 0. The areas considered in this analysis range from 1x1,3X3, 5X5, ..., 15X15, and the tracts from tract 0, tract 1, tract 2, ..., to tract 7. Since a township covers 6X6 mile^2, large areas and tracts defined above might consist of sections from more than one township. Each of the included sections must be referenced with a township/meridian range/section code, in order to query the PUR table to obtain methyl bromide applications in the included section by date. In Figure 4 the numbers inside the township are section numbers. A Perl program (township.pl) was developed to generate MTRSs for sections in an area (Appendix 1). Township.pl takes a station's STR as shown in Table 1, and prints on the screen or to a file a matrix of MTRSs for the square block of surrounding sections depending on the specified size.

Three meridians are used in California: Mount Diablo, San Bernardino and Humboldt. The Mount Diablo meridian covers the biggest area. All of the sampling sites are located in the Mount Diablo meridian [5]. However, one monitoring site in Kern County (MET) was very close to the boundary between the Mount Diablo and the San Bernardino meridians. Areas and tracts included sections in both systems. The program cannot handle this situation yet. In addition, the size and arrangement of sections at this boundary is not confined to 1x1 mile^2 section configuration used elsewhere. Therefore, this site was dropped from the analysis.

The emission of methyl bromide from soil could last up to several days, or could largely occur in the first 48 hours, depending on the application methods, soil status and meteorological conditions. Air sampling was taken from Monday through Thursday. Therefore, the use week relevant to a weekly average concentration was defined from Friday of the previous week to Thursday of the current monitoring week. The weekly use of methyl bromide over various areas was calculated with a Perl program (mb_use01.pl), which is appended to this document (Appendix II). The application date in the use report was used as a single date for the entire application. However, occasionally applications to a single field may span several days, but are only reported with a single application date. In these cases, it is very difficult to obtain multi-date application information.

2.4 Methods to relate the air concentration to the methyl bromide use

According to the Gaussian equation, air concentration is proportional to the flux rate under fixed soil status and weather conditions. When considering a large area and over a long period, this linear proportionality can be extended to the relationship between air concentration and the amount of methyl bromide used in the area. The Linear Regression Model was used to relate the air concentration to the methyl bromide use:

$$Y = a + bX \tag{1}$$

where Y is the average air concentration over a certain period (1 week, 3 to 4 weeks and 7 to 8 weeks), and X is the weekly average methyl bromide use over various areas or tracts in that period.

R^2 and Error Mean Square(EMS) measure the fitness of the Linear Regression Model. R^2 represents the percentage of variation of the dependent variable that is explained by the independent variable, and it is often referred to as the coefficient of determination. EMS is the average squared residual error not explained by the model, which is defined as:

$$EMS = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{n-2} \tag{2}$$

where n is the number of samples, Y_i and $\hat{Y}_i$ are measured and corresponding regression-estimated air concentrations, respectively. Higher R^2 and lower EMS characterize better regressions.

The least squares method was used to estimate regression coefficients a and b. Confidence intervals for a, b and R^2 were calculated using methods described in [6]. A computer program (linear.pl) was developed to conduct regression analysis (Appendix III).

If the regression analysis yields useful relationships, then given an air concentration C, the corresponding use, represented in the X variable, can be solved for by using the equation below:

$$X = (C - a)/b \tag{3}$$

The use, X, is in lb/week over certain areas (3x3, 5x5, ..., 15x15), and C, the concentration, is in ppb.

A randomization procedure was utilized to verify significance levels in the regression analysis [7]. This procedure was a computer intensive study conducted for each area regression (TESTC.FOR, Appendix IV). The eleven use values were randomly shuffled in relation to the air concentrations, a conventional regression was performed, and the resulting F ratio statistic was recorded. For each area regression data set, this procedure was repeated 10000 times. The resulting column of 10000 F ratios was sorted, and the original F ratio was compared to the distribution.

3. Results

3.1 Air concentration

Weekly average air concentrations (ppb) at the eleven sites are shown in Table 2. Based on weekly average concentrations, air concentrations over a longer period such as 3 or 4 weeks and 7 or 8 weeks were also calculated (Table 3).

The air concentration of methyl bromide changed from site to site and from week to week over the monitoring periods (Fig. 5). The highest concentration (15.58 ppb) was observed at PMS in Monterey County in week 5. In fact, the air concentration at PMS was consistently higher than other Monterrey stations except in week 1. In Kern County, CRS had higher concentrations than other sites. Moreover, air concentrations in all places appeared to be lower in some weeks, and higher in other weeks. For example, during weeks 4, 5 and 6, all sites in Kern County reported low air concentrations, while for Monterey/Santa Cruz Counties, low concentrations occurred in weeks 4 and 8.

Table 2: Weekly average air concentrations (ppb)

County	Site	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7	Week 8
Kern	ARB	0.507	0.132	0.292	0.111	0.039	0.059	0.188	
Kern	CRS	2.828	3.647	4.595	0.459	0.150	0.641	2.790	
Kern	MET	0.064	0.111	0.115	0.070	0.030	0.059	0.145	
Kern	MVS	0.061	0.095	0.066	0.096	0.034	0.092	0.201	
Kern	SHA	1.775	0.705	0.815	0.332	0.043	0.347	1.536	
Kern	VSD	0.112	0.099	0.091	0.104	0.033	0.081	0.175	
Monterey	CHU	0.730	1.300	0.340	0.400	0.260	1.610	0.590	0.110
Monterey	LJE	10.630	8.470	1.270	1.350	0.830	5.630	2.580	0.250
Monterey	OAS	0.380	0.440	0.170	0.400	0.250	1.010	0.390	0.080
Monterey	SAL	1.640	2.360	0.770	0.500	0.700	3.010	1.200	0.140
Monterey	PMS	5.170	8.140	9.890	1.270	15.580	9.490	11.210	1.110
Santa Cruz	SES	8.340	2.880	1.960	1.020	0.840	3.630	2.010	0.220

Note: The 1-week reference concentrations for adult and children are 120 ppb and 70 ppb respectively [4].

Table 3: Average air concentrations(ppb) over 3/4 weeks and 7/8 weeks

County	Site	Average air concentration (ppb)		
		1st 4-weeks	2nd 4-weeks or 3-weeks	7/8-weeks
Kern	ARB	0.261	0.095	0.19
Kern	CRS	2.882	1.194	2.16
Kern	MET	0.090	0.078	0.08
Kern	MVS	0.080	0.109	0.09
Kern	SHA	0.907	0.642	0.79
Kern	VSD	0.102	0.096	0.10
Monterey	CHU	0.693	0.643	0.67
Monterey	LJE	5.430	2.323	3.88
Monterey	OAS	0.348	0.433	0.39
Monterey	SAL	1.318	1.263	1.29
Monterey	PMS	6.118	9.348	7.73
Santa Cruz	SES	3.550	1.675	2.61

Note: The 8-week reference concentrations for adult and children are 2 ppb and 1 ppb respectively [4].

Many factors might have contributed to these highs and lows, such as weather conditions, methyl bromide use patterns, and topographical characteristics near the monitoring sites. In Figures 6 and 7, the weekly air concentration was compared to the weekly methyl bromide use in an area of 13x13. In Kern County, low methyl bromide use corresponded to low concentration except in week 5 at CRS (Fig. 7). Use data from the 13x13 area could not completely explain the weekly variation of air concentrations. The 13x13 use data of Monterey/Santa Cruz Counties also show a correlation between weekly air concentration and weekly use, but fail to explain every single data point. A single application in the use report listed on a single date, but actually occurring on several dates, would introduce greater variability into the weekly concentration/use comparisons than for the longer time period comparisons.

3.2 Effect of Temporal Scales

In the regression model (1), the relationship between air concentration (Y) and methyl bromide use (X) was determined over various time periods. By examining R^2 and EMS values, we can determine over what period the methyl bromide use is most closely related to the air concentration.

Table 4 shows correlations between average air concentration and methyl bromide use in various areas and tracts and over various periods. The correlation coefficient between air concentration and methyl bromide use is significant over many areas and time periods (Table 4). R^2 values are higher over longer periods. However, the significant R^2-value thresholds also increase when the number of samples decreases. For most areas and tracts, the EMS declined with longer periods. The regression model using 7 to 8-week

Table 4: R^2 between average air concentration (ppb) and average methyl bromide usage (lb/week) over various areas, tracts and periods

Area/Tract	Time period over which the average value was calculated					
	1 week (n = 83)		3/4 weeks (n = 22)		7/8 weeks (n = 11)	
	R^2	EMS	R^2	EMS	R^2	EMS
area 1x1	0.211	7.53	0.287	4.27	0.898	0.60
area 3x3	0.646	3.38	0.806	1.16	0.917	0.49
area 5x5	0.500	4.77	0.683	1.90	0.912	0.52
area 7x7	0.575	4.06	0.699	1.80	0.946	0.32
area 9x9	0.519	4.59	0.645	2.17	0.871	0.76
area 11x11	0.434	5.41	0.580	2.51	0.742	1.52
area 13x13	0.387	5.85	0.558	2.65	0.695	1.81
area 15x15	0.380	5.91	0.542	2.74	0.667	1.97
tract 0	0.211	7.53	0.287	4.27	0.898	0.60
tract 1	0.612	3.70	0.840	0.96	0.890	0.65
tract 2	0.225	7.40	0.416	3.50	0.705	1.74
tract 3	0.471	5.05	0.625	2.24	0.845	0.92
tract 4	0.166	7.96	0.302	4.18	0.485	3.04
tract 5	0.069	8.88	0.212	4.72	0.217	4.63
tract 6	0.030	9.25	0.143	5.13	0.140	5.09
tract 7	0.094	8.65	0.174	4.94	0.174	4.88
Significant R^2 values						
$R^2_{0.05}$	0.047		0.179		0.362	
$R^2_{0.01}$	0.080		0.288		0.540	

Note: R^2 is often referred as the coefficient of determination, representing the percentage of variation of air concentration that is explained by the use of methyl bromide. EMS is the average squared residual error not explained by the model.

average data generated the lowest EMS values. Increased noise in the concentration-methyl bromide use relationship was filtered out when the averaging period lengthened.

Because the 7 to 8-week averaging period yielded the highest correlation and the lowest EMS values, and because the main concern of this study is subchronic effects on the order of 6-8 weeks exposure, analyses in the following paragraphs will be based on the 7 to 8-week average data.

3.3 Effect of spatial scales

Dispersion of methyl bromide may reach several miles away from the application sites. However, methyl bromide use in a certain sized area or at a certain distance around the monitoring sites might have a better correlation to the air concentration. The methyl bromide uses in various areas and tracts around each monitoring site, along with the air concentration are shown in Table 5 and Figures 8 and 9.

Table 5: Monitored average air concentrations and reported average weekly methyl bromide uses in Kern, Monterey and Santa Cruz counties over the period of 7 or 8 weeks

County	**Site**	**Average concentration (ppb)**	**Mean of weekly methyl bromide use (lb/week)**							
			1X1	**3X3**	**5X5**	**7X7**	**9X9**	**11X11**	**13X13**	**15X15**
Kern	ARB	0.19	0	0	0	0	0	45	45	45
Kern	CRS	2.16	0	955	9448	9448	9448	9448	15085	15085
Kern	MVS	0.09	0	0	0	0	0	0	32	77
Kern	SHA	0.79	0	0	0	955	9448	9448	9448	9448
Kern	VSD	0.10	0	0	0	0	0	0	0	32
Monterey	CHU	0.67	0	1202	1202	5360	7883	12444	15384	16886
Monterey	LJE	3.88	1114	6259	8590	23630	31427	40985	56066	62146
Monterey	OAS	0.39	0	0	591	2306	2306	2306	2306	2803
Monterey	SAL	1.29	0	0	4352	14848	26821	46393	59637	63250
Monterey	PMS	7.73	2232	13633	26326	51372	66951	77251	80595	82312
Santa Cruz	SES	2.61	169	6629	14111	24311	39339	55178	60308	66495

Methyl bromide use increased as area increased. The air concentration agreed well with the methyl bromide use amount in both low and high ends. All area regressions were highly significant ($p<.01$). The correlation between air concentration and use over the 7X7 area showed the best results in terms of the highest R^2 and lowest EMS (Table 4). The correlation between the air concentration and the methyl bromide use tracts was highest for tracts 0, 1, 2 and 3, and generally tapered off from tract 4 and above as the tracts became further away from the central section (Table 4).

Linear regression between the air concentration and methyl bromide use was conducted over various areas (Figures 10 and 11, Table 6). In the linear model (1), Y is air concentration (ppb), and X is the methyl bromide use in an area (lb/week). Regression coefficients ‘a’ and ‘b’ have a clear meaning: a represents the air concentration when there is zero pounds of methyl bromide used in the considered area during the 7/8 week period, and b represents the increase of air concentration resulting from one unit increased methyl bromide use. The coefficient ‘a’ could also be interpreted as the background concentration for that given area since it represents the estimated concentration when zero pounds of methyl bromide are applied in the use area that was defined for any particular regression. However, only the additive constant for the 1x1 area was significantly different from zero. While none of the remaining additive constant estimates of ‘a’ were significantly different from zero, there was a general decrease in magnitude, as the size of the base area increased. This suggested that as the area size increased, the leftover background concentration decreased.

The coefficient b decreased exponentially when the area increased (Fig. 12). This is a natural consequence of greater use in the larger areas.

Table 6: Regression coefficients, R^2, and EMS values for different area regressions

area	a	b	R^2	EMS
1x1	0.833	0.003054	0.898	0.60
3x3	0.510	0.000498	0.917	0.49
5x5	0.258	0.000264	0.912	0.52
7x7	0.118	0.000141	0.946	0.32
9x9	0.037	0.000101	0.871	0.76
11x11	0.118	0.000073	0.742	1.52
13x13	0.093	0.000063	0.695	1.81
15x15	0.121	0.000058	0.667	1.97

A good fit by the linear regression is indicated by a higher R^2 and lower EMS. R^2 increased in going from small to mid-sized areas. It peaked at the 7x7 area. At 9x9 and above, it declined. EMS showed the corresponding pattern. The best fit was obtained from use data over the area of 7x7 (Table 6).

Although the regression coefficients and correlation coefficients differ with the size of areas, statistically, the regression lines are probably not different from each other. One difficulty in comparing values between different regressions is that for the area analysis, the values are not independent. Each successively larger area contains the previous areas. The highest R^2 value occurred with 7x7. This area includes tracts 0, 1, 2, and 3. Tracts 0,1, 2 and 3 showed the highest correlations with the concentrations in the monitoring section. However, the R^2 value for tract 4 declined (Table 4), a trend which continued with the larger and more distant tracts. The general decline in ‘a’ and ‘b’ values, together with the peaking of the R^2 value at 7x7suggest that the 7x7 area use is a reasonable choice in explaining the variability of 8 week average air concentrations.

Table 7: 95% confidence Intervals for a, b and R^2

Area	a			b			R^2		
	estimate	CI_1	CI_2	estimate	CI_1	CI_2	estimate	CI_1	CI_2
1x1	0.833	0.249	1.418	0.003054	0.002278	0.003829	0.90	0.65	0.97
3x3	0.510	-0.050	1.070	0.000498	0.000386	0.000611	0.92	0.71	0.98
5x5	0.258	-0.353	0.871	0.000264	0.000202	0.000326	0.91	0.69	0.98
7x7	0.118	-0.373	0.610	0.000141	0.000115	0.000166	0.95	0.80	0.99
9x9	0.037	-0.749	0.824	0.000101	0.000071	0.000130	0.87	0.57	0.97
11x11	0.118	-1.009	1.246	0.000073	0.000041	0.000106	0.74	0.29	0.93
13x13	0.093	-1.162	1.349	0.000063	0.000032	0.000095	0.69	0.22	0.91
15x15	0.121	-1.193	1.436	0.000058	0.000027	0.000089	0.67	0.18	0.90

3.4 Sensitivity

The highest 7/8 week monitoring results occurred at PMS in Monterey. There is a fumigation chamber approximately 600 feet to the east of the PMS monitoring location which utilizes an estimated 100 lbs of methyl bromide per week, or 800 lbs during the 8 week monitoring period in Monterey (Monterey Agricultural Commissioner Staff, personal communication). This mass compares with the 8x2,232 pounds/week=17,856 pounds used within 1 mile during the 8 week period or 8x16,333 pounds/week=130,664 pounds used within 2 miles of the monitoring site during the same period. The mass utilized in the commodity chamber is small relative to the mass applied to the soil in the nearby area.

To further examine the possible effect of the nearby fumigation chamber, the regression equation for the 7x7 area was recalculated after omitting the PMS data point. The results were similar to the original results with all of the data points. The additive constant changed from 0.118 to 0.240, while the multiplicative constant changed from 0.000141 to 0.000121. The R^2 value lowered from 0.95 to 0.84. The regression remained highly significant ($p<.01$).

For each of the area regressions, the randomization test [7] resulted in statistical significance ($p<.01$). This randomization procedure does not assume normality, nor make any assumptions about correlations between locations.

These sensitivity analyses did not change the main relationship between use and concentrations measured over the 7/8 week period. High use areas exhibited higher measured air concentrations.

3.5 Air Concentrations Estimated From Pesticide Use

As discussed previously, air concentrations can be estimated from the amount of methyl bromide with equation (3). Analysis of the data indicates that methyl bromide use averaged for seven to eight weeks within a 7x7 mile^2 area gives the best correlation to air concentration. Therefore, the following equation gives the air concentration as a function of weekly pesticide use over a 7x7 $mile^2$ area.

$$\text{Air Concentration (ppb)} = 0.118 + 0.000141 * \text{Methyl Bromide (lbs /[week *7x7miles}^2\text{])}$$

Using the subchronic reference concentration of 1 ppb as a benchmark, the corresponding use amount was calculated. The result for the full data set was 6255 lbs/[week *7x7$miles^2$]. This use amount corresponds to 7x7 area, or 49 square miles. A more convenient unit for regulatory purposes is pounds of methyl bromide per month (30 days) in a township (36 square miles). Adjusting for the time area and area differences, 6255 lbs/[week*7x7$miles^2$] in 49 square miles is equivalent to 19,695 lbs/month in a township, using 1 ppb as the benchmark. The same calculation, using the regression which omitted the monitoring site with the fumigation chamber nearby (PMS), yielded 19,776 lbs/month. For regulatory purposes, this number should be rounded to 20,000 lbs/month per township.

The earlier draft of this analysis described 18,000 pounds per township per month as equivalent to 1 ppb. The change to 20,000 pounds is due to correction of some errors in the earlier draft. As discussed earlier, the location of several monitoring stations were incorrectly described in ARB's monitoring report.

3.6 Areas of High Air Concentrations

The highest subchronic exposures are expected to occur in townships where 20,000 pounds is exceeded each month for three consecutive months or more. The 1999 and 2000 pesticide use reports were examined to determine the areas where these amounts are exceeded. (The year 2000 use reporting data used for this analysis must be considered very preliminary and is subject to change after error-checking programs are run.) In 1999, 15.4 million pounds of methyl bromide were applied in 458 townships in 45 of the 58 California counties (Figure 13). Seventeen townships in eight counties exceeded 20,000 pounds each month for three consecutive months or more in 1999 (Table 8, Figures 13 - 21). Use decreased significantly in 2000, with 10.3 million pounds of methyl bromide applied in 415 townships of 41 counties (Figure 22). Thirteen townships in seven counties exceeded 20,000 pounds each month for three consecutive months in 2000 (Table 9, Figures 22 - 30).

Table 8: Pounds of methyl bromide by month in townships exceeding 20,000 pounds in three consecutive months for 1999

Township	County	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec	Total
M25S26E	KERN									23,500	28,295	29,671		81,466
M27S25E	KERN		9,800			35,319		9,643	43,171	159,697	65,491		453	323,573
M12S02E	**MONTEREY**	**259**			**4,763**	**6,639**	**4,498**	**33,160**	**87,704**	**176,507**	**181,006**	**11,210**		**505,745**
M13S02E	MONTEREY						2,883	6,046	47,872	98,922	120,166	15,237		291,126
M13S03E	MONTEREY							15,205	32,460	46,750	55,018	8,577	1,318	159,328
M14S03E	MONTEREY				1,603		31,499	87,555	67,539	114,967	29,283	6,283		338,728
S06S08W	ORANGE		7,973	5,930				21,092	57,147	36,474	1,733			130,348
S07S09E	RIVERSIDE	77,900	7,193									30,968	73,570	189,632
S09N33W	SANTA BARBARA			1,876		13,775				27,488	41,256	24,390		108,784
S10N34W	SANTA BARBARA		1,134			2,063			56,044	282,827	170,432	3,994	3,285	519,778
M11S02E	SANTA CRUZ	1,367			9,474	9,862		310	60,283	97,288	119,359	23,245		321,189
M12S01E	SANTA CRUZ				2,640	3,381	16,893	21,199	36,981	45,873	34,769	12,925		174,660
M12S02E	**SANTA CRUZ**								**27,668**	**33,348**	**49,230**	**4,689**		**114,936**
M14N03E	SUTTER	4,752			121			3,722	48,720	32,195	36,070	21,766	5,066	152,411
S01N21W	VENTURA		1,200	2,588	26,095	49,741	25,249	29,699	153,274	262,675	1,254	199	955	552,928
S01N22W	VENTURA	687	1,642	12,545	920	48,288	67,996	32,842	61,511	33,422	3,393	1,276	1,235	265,755
S02N21W	VENTURA	4,980		9,555	12,458	15,267	4,230	95,063	84,558	59,058		3,078	225	288,470
S02N22W	VENTURA	3,136	12,499	2,964	12,653	36,831	101,343	122,488	243,118	96,642	12,493	9,209	24,225	677,601
	Subtotal	93,082	41,440	35,457	70,726	221,166	254,590	478,024	1,108,049	1,627,630	949,247	206,716	110,331	5,196,458
	Statewide Total	852,210	474,660	728,293	736,839	669,149	599,053	963,993	2,140,574	3,103,346	2,572,225	1,514,538	1,025,232	15,380,691
	Percent	11	9	5	10	33	42	50	52	52	37	14	11	34

Table 9: Pounds of methyl bromide by month in townships exceeding 20,000 pounds in three consecutive months for 2000

Township	County	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
M27S25E	KERN			5,695		31,211	23,419	70,393	25,915	8,149		1,184		165,967
M07S11E	MERCED	50,607	87,242	40,554	51,774	8,678							25,450	264,305
M12S02E	**MONTEREY**			**1,233**	**2,035**	**9,159**	**21,324**	**5,546**	**61,650**	**149,037**	**99,604**	**8,525**	**804**	**358,918**
M13S02E	MONTEREY					1,867	14,466	5,159	51,416	79,043	88,833	19,081		259,865
M13S03E	MONTEREY								4,029	58,040	39,720	21,291		123,080
M14S03E	MONTEREY					23,664	48,373	40,439	43,784	90,127	34,506	774		281,666
S06S08W	ORANGE			1,039	500		16,591	21,742	28,163	26,776				94,809
S10N33W	SANTA BARBARA		2,250	1,575			858		21,825	86,883	37,639			151,030
M11S02E	SANTA CRUZ	274		2,486	12,010	2,620		9,363	38,763	82,322	76,281	5,626		229,744
M12S01E	SANTA CRUZ			2,165	4,442	5,809	14,793	28,382	43,148	54,187	20,399	8,783		182,108
M12S02E	**SANTA CRUZ**								**23,657**	**55,160**	**30,711**			**109,528**
S01N21W	VENTURA	681		2,866	17,955	14,185	7,739	66,077	176,720	41,847	3,825	7,136	1,013	340,043
S02N21W	VENTURA			2,728	24,053	7,169		20,765	109,625	47,674	1,444		51	213,509
S02N22W	VENTURA	11,708	450	3,786	9,485	14,805	52,907	201,375	149,593	82,298	4,604	4,565	4,570	540,145
	Subtotal	63,270	89,942	64,128	122,254	119,166	200,469	469,240	778,287	861,545	437,564	76,964	31,887	3,314,718
	Statewide Total	564,150	547,721	768,057	603,817	429,341	490,338	820,941	1,677,722	1,845,954	1,553,363	487,694	557,337	10,346,436
	Percent	11	16	8	20	28	41	57	46	47	28	16	6	32

4. Summary

This analysis examined the relationship between use, area and time period and measured air concentrations for methyl bromide. There were significant regression relationships between use and measured air concentrations for differing time periods and differing area sizes. All 7/8 week regressions were significant, and underscored a fundamental positive correlation between level of use of methyl bromide as a soil fumigant and measured air concentrations. The highest R^2 value relating use and concentration occurred with the 7x7 area over a period of 7/8 weeks .

There are several caveats to this analysis. First, this analysis only includes pesticide use data from field fumigations. Pesticide use data for structural, commodity, and other types of methyl bromide fumigations are not amenable to this type of analysis because they do not include information on specific location or date, and are incomplete for year 2000. Structural or commodity fumigations may have occurred during the monitoring, but there is no way to take their contribution to the air concentrations into account. However, these effects were probably minor, based on the strength of the statistical relationships determined in the analysis and the relative amount of pounds applied for structural/commodity fumigations and soil fumigations in historical data. Also, omitting the high value from the 7x7 regression did not make a substantial change to the regression result. Second, this analysis assumes that all pesticide use data for field fumigations is complete and accurate. Missing or incorrect data could significantly alter the regressions. Missing data would cause an underestimation of the amount of methyl bromide that correlates with a specific air concentration. Incorrect data, where reported use was inflated, would cause an overestimation of the amount of methyl bromide that correlates with a specific air concentration. We think overestimates in reported methyl bromide use are unlikely, however, since use rates were within reasonable ranges. Third, the regression line represents the mean estimate of concentration. For a given use amount, one would expect the corresponding concentration to be greater than the mean estimate about half of the time and less than the mean estimate about half of the time. Fourth, while there are significant differences in emission rates between methods over a 24-hour period, it is likely that there is little difference between methods in emissions over several weeks. Adjustments for method differences do not appear to be necessary for subchronic exposure mitigation. However, additional monitoring is needed to verify this assumption.

Acknowledgment

We appreciate the help from Craig Nordmark and Johanna Walters, who provided GIS maps for locating monitoring sites for this project.

References

[1]John Sanders, June 16, 2000. Memorandum from John Sanders to George Lew. Subject: Recommendation for 1,3-dichloropropene and methyl bromide monitoring for the toxic air contaminant program

[2] ARB, 2001. Ambient air monitoring for methyl bromide and 1,3-Dichloropropene in Monterrey/Santa Cruz Counties - Fall 2000. California Air Resources Board. Sacramento, CA.

[3] ARB, 2001. Ambient air monitoring for methyl bromide and 1,3-Dichloropropene in Kern County - Summer 2000. California Air Resources Board. Sacramento, CA.

[4]DPR, 1999. Methyl bromide risk characterization document for inhalation exposure (Draft RCD 99-02). California Department of Pesticide Regulation, Sacramento, CA.

[5] DWR. Undated. Numbering water wells in California. California Department of Water Resources. Sacramento, CA.

[6]Agresti Finlay, 1986. Statistical Methods for the Social Sciences, 2nd Edition, Dellen Macmillan. P253-273.

[7]Sokal, Robert R. and F. James Rohlf. 1981. Biometry. W.H. Freeman and Company, New York, New York. p.791-795

FIGURES

Appendix I.

A Perl program for determining the neighboring sections in the township & meridian range system

1. Important notes

- Two coordination systems separately for the township&Range, and sections
- For township&Range, must have algorithm
    to calculate the numerical component, and
    to determine the directional component
- For sections, must have algorithm
    to convert from XY coordinates to section number, and
    to convert from section number to XY coordinates
- Three systems are used in California, Boundaries issues between two systems (not solved yet)

- two types of notation: MTRS and STR, the algorithm must be able to parse and assembly the two notations
    -The input is in STR format, which was used in ARB's reports
    -The output is in MTRS, which was adopted in PUR reports

2. Source file (township.pl)

```
#!/usr/local/bin/perl -w
#    Last change:  LI   12 Apr 2001    9:57 am
# township.pl
# Generates strings representing the surrounding MTRS(Meridian Township Range
Section)
# for a giving MTRS within the specified distance
# -----------------------------------------------------------

# all parameters are fed from command line
# usage: township.pl MTRS DX,DY
# Please note two forms of MTRS: S.3/T.16S/R.3E or M16S03E03
# The first form was used for arb monitoring project
# The second for was used in DPR's PUR report
# This program assumes the input in the first format and generates MTRS in the
2nd format



#my ($MTRS, $DX, $DY)= @ARGV;
#print "@ARGV\n";

$working_dir = 'E:\Arb\1807';
chdir $working_dir;

sub_MTRS("S.19/T.31S/R.29E", 3,7);
```

```
# ----------------------------------------------------------- #
# usage: sub_MTRS(STR, DX, DY)
sub sub_MTRS {
my ($STR, $DX, $DY) = @_;

# extracting section number, township and range for the sampling site
$STR =~ m/S.(\d+)\/T.(\d+)([A-Z])\/R.(\d+)([A-Z])/;
$S=$1;
$T_val = $2;
$T_dir = $3;
$R_val = $4;
$R_dir = $5;
$MTRS= "M"."$T_val"."$T_dir"."$R_val"."$R_dir"."$S";
print "$STR\t $MTRS\n";
#print "$S,  $T_val,  $T_dir,  $R_val,  $R_dir\n";

# get coordinate for the sampling section
 $n = int($S/6) + 1;
 if($n==1 or $n==3 or $n==5) {
    $m = $n*6 + 1 - $S;}
 else {
    $m = $S - ($n-1)*6;}
 #print "$m, $n\n";

# calculating MTRS for surrounding grids


for ($j=-$DY; $j<=$DY; $j++){
for ($i=-$DX; $i<=$DX; $i++){
        # first, get coordinate for the surrounding sections, also the
township value and range value
        $sx =$m+$i;
        if($sx>6){
            $RR_val = $R_val + int($sx/6);
            $sx=$sx-6*int($sx/6);}
        elsif($sx<1){
             $RR_val = $R_val - (1+abs(int($sx/6)));
             $sx=$sx+6*(1+abs(int($sx/6)));}
        else {
             $RR_val = $R_val;}

        $sy =$n+$j;
        if($sy>6){
            $TT_val = $T_val + int($sy/6);
            $sy=$sy-6*int($sy/6);}
        elsif($sy<1){
             $TT_val = $T_val - (1+abs(int($sy/6)));
             $sy=$sy+6* (1+abs(int($sy/6)));}
        else {
             $TT_val = $T_val;}

        # The directions for township and range are the same with those of
sampling site
        # Need more analysis here
        $TT_dir = $T_dir;
        $RR_dir = $R_dir;

        # then, get the section number from its xy coordinates
```

```
        if($sy==1 or $sy==3 or $sy==5) {
           $SS = $sy*6 - $sx + 1;}
        else {
           $SS = ($sy-1)*6 + $sx;}

        # now, get the MTRS
        if ( $TT_val <=9 ) {$TT_val = "0"."$TT_val";}
        if ( $RR_val <=9 ) {$RR_val = "0"."$RR_val";}
        if ( $SS <=9 ) {$SS = "0"."$SS";}

        $new_MTRS = "M"."$TT_val"."$TT_dir"."$RR_val"."$RR_dir"."$SS";
        print "$new_MTRS ";

} # end of i loop
print "\n";
} # end of j loop

} # end of sub
```

3. Examples

Surrounding sections of a monitoring site (PMS 7X7)

```
Command Prompt

E:\ARB\1807>township.pl
S.9/T.12S/R.2E    M12S2E9
M11S01E25 M11S02E30 M11S02E29 M11S02E28 M11S02E27 M11S02E26 M11S02E25
M11S01E36 M11S02E31 M11S02E32 M11S02E33 M11S02E34 M11S02E35 M11S02E36
M12S01E01 M12S02E06 M12S02E05 M12S02E04 M12S02E03 M12S02E02 M12S02E01
M12S01E12 M12S02E07 M12S02E08 M12S02E09 M12S02E10 M12S02E11 M12S02E12
M12S01E13 M12S02E18 M12S02E17 M12S02E16 M12S02E15 M12S02E14 M12S02E13
M12S01E24 M12S02E19 M12S02E20 M12S02E21 M12S02E22 M12S02E23 M12S02E24
M12S01E25 M12S02E30 M12S02E29 M12S02E28 M12S02E27 M12S02E26 M12S02E25
```

Surrounding sections of a monitoring site (VSD 7X15)

```
Command Prompt

E:\ARB\1807>township.pl
S.19/T.31S/R.29E          M31S29E19
M30S28E15 M30S28E14 M30S28E13 M30S29E18 M30S29E17 M30S29E16 M30S29E15
M30S28E22 M30S28E23 M30S28E24 M30S29E19 M30S29E20 M30S29E21 M30S29E22
M30S28E27 M30S28E26 M30S28E25 M30S29E30 M30S29E29 M30S29E28 M30S29E27
M30S28E34 M30S28E35 M30S28E36 M30S29E31 M30S29E32 M30S29E33 M30S29E34
M31S28E03 M31S28E02 M31S28E01 M31S29E06 M31S29E05 M31S29E04 M31S29E03
M31S28E10 M31S28E11 M31S28E12 M31S29E07 M31S29E08 M31S29E09 M31S29E10
M31S28E15 M31S28E14 M31S28E13 M31S29E18 M31S29E17 M31S29E16 M31S29E15
M31S28E22 M31S28E23 M31S28E24 M31S29E19 M31S29E20 M31S29E21 M31S29E22
M31S28E27 M31S28E26 M31S28E25 M31S29E30 M31S29E29 M31S29E28 M31S29E27
M31S28E34 M31S28E35 M31S28E36 M31S29E31 M31S29E32 M31S29E33 M31S29E34
M32S28E03 M32S28E02 M32S28E01 M32S29E06 M32S29E05 M32S29E04 M32S29E03
M32S28E10 M32S28E11 M32S28E12 M32S29E07 M32S29E08 M32S29E09 M32S29E10
M32S28E15 M32S28E14 M32S28E13 M32S29E18 M32S29E17 M32S29E16 M32S29E15
M32S28E22 M32S28E23 M32S28E24 M32S29E19 M32S29E20 M32S29E21 M32S29E22
M32S28E27 M32S28E26 M32S28E25 M32S29E30 M32S29E29 M32S29E28 M32S29E27
```

4. References

DWR. Undated. Numbering water wells in California. California Department of Water Resources. Sacramento, CA.

Appendix II.

A Perl program for calculating weekly zone use of MeBr for each monitoring site

1. Notes

The program calculates the weekly zone use of MeBr in various areas (3x3 5x5, 7x7 ... 15x15).

The program calls the subroutine (township.pl) described in Appendix I**.**

2. Source file (mb_use01.pl)

```
#!/usr/local/bin/perl

#    Last change:  LI    1 May 2001    2:02 pm
# mb_use01.pl
# calculates total amount of methyl bromide use in surrounding area of
monitoring sites
# ----------------------------------------------------------

# all parameters are fed from command line
# usage: township.pl MTRS DX,DY
# Please note two forms of MTRS: S.3/T.16S/R.3E or M16S03E03
# The first form was used for arb mornitoring project
# The second for was used in DPR's PUR report
# This program assumes the input in the first format and generates MTRS in the
2nd format



#my ($MTRS, $DX, $DY)= @ARGV;
#print "@ARGV\n";

$working_dir = 'E:\Arb\1807';
chdir $working_dir or die "couldn't find the path $working_dir\n";

$infile1='station.dat';
$infile2='weekly_con.dat';
$infile3='PUR.dat';
#$infile3='PUR_updated.dat';

#$outfile0='weekly_con_use0.dat';
$outfile1='weekly_con_use1.dat';
$outfile2='weekly_con_use2.dat';
$outfile3='weekly_con_use3.dat';
$outfile4='weekly_con_use4.dat';


open IN1, "$infile1";
```

```
open OUT1, ">$outfile1";
open OUT2, ">$outfile2";
open OUT3, ">$outfile3";
open OUT4, ">$outfile4";
open PMS, '>PMS_test.txt';

print OUT1 "ID   County  Station Week    Conc    Appl(1) Appl(2) Appl(3)
Appl(4) Appl(5) Appl(6) Appl(7)\n";
print OUT2 "ID   County  Station Week    Conc    Appl(1) Appl(2) Appl(3)
Appl(4) Appl(5) Appl(6) Appl(7)\n";
print OUT3 "ID   County  Station Week    Conc    Appl(1) Appl(2) Appl(3)
Appl(4) Appl(5) Appl(6) Appl(7)\n";
print OUT4 "ID   County  Station Week    Conc    Appl(1) Appl(2) Appl(3)
Appl(4) Appl(5) Appl(6) Appl(7)\n";


L1: while ($line_IN1 = <IN1>) {
#get the station record
chomp $line_IN1;
($id, $county, $station, $STR1,$total) = split(/\,/,$line_IN1);
print "$id, $county, $station, $STR1, $total\n";



  # skip the title line
  if ($id eq 'ID') {
    next L1;}

  # determine the starting date, number of weeks for air sampling,
  # and the number of days in the first week,

  if ($county eq "Kern"){# for most weeks, sampling usually started from Mon
and ended at Thur
    $num_wk=7;           # However, the first week in Kern county started on
Wendesday, and ended on Thursday.
                         # we assuming that the average of concentration of
these two days represents the average of that week
    $starting_day=201-5; # 201 is the first sampling date in Kern county
(7/19, Wed),
    $days_wk1=2; }       # the julian day for the previous Friday was 201-5.
                         # the day should be shift back 5 days
  else {$num_wk=8;       # In Monterey county, the first sampling date was
Monday(day 255, 09/11).

    $starting_day=255-3; # this number should correspond to 09/08/00, the
first Friday before monitoring started
    $days_wk1=4;}
  $ending_day = $starting_day + 6;

  #get the weekly concentration record
  open (IN2, "$infile2") or die "could not open file $infile2!\n";
   L2: while ($line_IN2 = <IN2>) {
      chomp $line_IN2;
      @wkc= split(/\t/,$line_IN2);
      print "@wkc\n";
      if ($wkc[0] eq $station) {
      last L2;}
     } # end of L2 loop
```

```
  close (IN2);

  # get the weekly application within the specified distances
  for ( $wk=1;$wk<=$num_wk;$wk++)
    {

      for ( $dist=1;$dist<=7 ;$dist++ )
        {
           ($section_ref,$distance_ref) = sub_MTRS($STR1,$dist,$dist);
           @section = @{$section_ref};
           @distance = @{$distance_ref};

           #print "@section\n";
           $num_sec=@section;
           print "$station, $wkc[0], $wk, $dist, $num_sec\n";

           chdir $working_dir or die "couldn't find the path $working_dir\n";

           open (IN3, "$infile3") or die"couldn't open file $infile3\n";

           $weekly_use1[$dist]=0;
           $weekly_use2[$dist]=0;
           $weekly_use3[$dist]=0;
           $weekly_use4[$dist]=0;

           while ($line_IN3 = <IN3>)
            {
             chomp $line_IN3;
             @use = split(/\t/,$line_IN3);
             #print "$use[2], $use[6]\n";

             for ( $l=0;$l<=$num_sec-1;$l++ )
                 {
                    #if ($use[2] eq $section[$l]){
                    if ( ($use[2] eq $section[$l]) and
($use[8]>=$starting_day) and ($use[8]<=$ending_day) ){
                       #$time_factor =
1/(abs($use[8]-($starting_day+$ending_day)/2)+.5);
                       #$dist_factor = 1/($distance[$l]**2);

                       if ($station eq "PMS" &&  $wk==8 && $dist<=3) {
                       print PMS "@use\n";}

                       $delt_t = $use[8]-($starting_day+$ending_day)/2 + 3;
                       #if ( $county eq "Kern" and $wk == 1) {
                       #   @t_factor = (0, 0, 0.1, 0.2, 0.4, 1, 0.85);}
                       #else {
                       #   @t_factor = (0.1, 0.2, 0.4, 1, 1, 0.85, 0.7);}
                       @t_factor = (0.1, 0.2, 0.4, 1, 1, 0.85, 0.7);
                       $time_factor = $t_factor[$delt_t];
                       #print "delt = $delt_t ; t_factor =$time_factor\n";

                       $dist_factor = 1/$distance[$l];

                       $weekly_use1[$dist] += $use[6];
         # no time and dist adjust
                       $weekly_use2[$dist] += $use[6] * $time_factor;
         # time adjust
```

```
                       $weekly_use3[$dist] += $use[6] * $dist_factor;
         # dist adjust
                       $weekly_use4[$dist] += $use[6] * $time_factor *
$dist_factor;     # dist adjust
                       }
                  } #end of for l

            } #end while IN3
            close IN3;
        }  # end for dist

     #print to OUT1 file
     print OUT1 "$id\t$county\t$station\t$wk\t$wkc[$wk]";
     for ($dist=1; $dist<=7; $dist++){
         $weekly_use1[$dist] = int($weekly_use1[$dist]+.5);
         print OUT1 "\t$weekly_use1[$dist]";}
     print OUT1 "\n";

     #print to OUT2 file
     print OUT2 "$id\t$county\t$station\t$wk\t$wkc[$wk]";
     for ($dist=1; $dist<=7; $dist++){
         $weekly_use2[$dist] = int($weekly_use2[$dist]+.5);
         print OUT2 "\t$weekly_use2[$dist]";}
     print OUT2 "\n";

     #print to OUT3 file
     print OUT3 "$id\t$county\t$station\t$wk\t$wkc[$wk]";
     for ($dist=1; $dist<=7; $dist++){
         $weekly_use3[$dist] = int($weekly_use3[$dist]+.5);
         print OUT3 "\t$weekly_use3[$dist]";}
     print OUT3 "\n";

     #print to OUT4 file
     print OUT4 "$id\t$county\t$station\t$wk\t$wkc[$wk]";
     for ($dist=1; $dist<=7; $dist++){
         $weekly_use4[$dist] = int($weekly_use4[$dist]+.5);
         print OUT4 "\t$weekly_use4[$dist]";}
     print OUT4 "\n";


  $starting_day+=7;
  $ending_day = $starting_day + 6;
  } # end for wk

} # end of L1 loop

close IN1, OUT1, OUT2, OUT3, OUT4, PMS;





# ----------------------------------------------------------- #
# usage: sub_MTRS(STR, DX, DY)
sub sub_MTRS {
my ($STR, $DX, $DY) = @_;
my ($MTRS,$LX,$LY, $m, $n, $i, $j, $sx, $sy);
```

```
my ($S, $T_val,$T_dir,$R_val, $R_dir);
my ($SS, $TT_val,$TT_dir,$RR_val, $RR_dir);
my (@new_MTRS);

# extracting section number, township and range for the sampling site
$STR =~ m/S.(\d+)\/T.(\d+)([A-Z])\/R.(\d+)([A-Z])/;
$S=$1;
$T_val = $2;
$T_dir = $3;
$R_val = $4;
$R_dir = $5;

# get the MTRS format for the sampling site
if ( $T_val <=9 ) {$T0_val = "0"."$T_val";} else { $T0_val = $T_val;}
if ( $R_val <=9 ) {$R0_val = "0"."$R_val";} else { $R0_val = $R_val;}
if ( $S <=9 ) {$S0 = "0"."$S";} else{$S0 = $S;}
$MTRS= "M"."$T0_val"."$T_dir"."$R0_val"."$R_dir"."$S0";

#print "$STR\t $MTRS\n";
#print "$S,  $T_val,  $T_dir,  $R_val,  $R_dir\n";

# create a file to store all of the neihbouring MTRS
$LX = $DX*2+1;
$LY = $DY*2+1;
if ( $LX<=9 ) {$LX="0"."$LX";}
if ( $LY<=9 ) {$LY="0"."$LY";}
$mtrs_out = "$MTRS"."($LX".X."$LY)".'.txt';

chdir "$working_dir/temp" or die "couldn't find the path
$working_dir\\temp\n";
open MTRS_OUT, ">$mtrs_out";
open DIST_OUT, ">$dist_out";

# get coordinate for the sampling section
 $n = int($S/6) + 1;
 if($n==1 or $n==3 or $n==5) {
    $m = $n*6 + 1 - $S;}
 else {
    $m = $S - ($n-1)*6;}
 # print "$m, $n\n";

# calculating MTRS for surrounding grids

$k=0;
for ($j=-$DY; $j<=$DY; $j++){
for ($i=-$DX; $i<=$DX; $i++){

        # first, get coordinate for the surrounding sections, also the
township value and range value
        $sx =$m+$i;
        if($sx>6){
            $RR_val = $R_val + int($sx/6);
            $sx=$sx-6*int($sx/6);}
        elsif($sx<1){
             $RR_val = $R_val - (1+abs(int($sx/6)));
             $sx=$sx+6*(1+abs(int($sx/6)));}
        else {
             $RR_val = $R_val;}
```

```
        $sy =$n+$j;
        if($sy>6){
            $TT_val = $T_val + int($sy/6);
            $sy=$sy-6*int($sy/6);}
        elsif($sy<1){
             $TT_val = $T_val - (1+abs(int($sy/6)));
             $sy=$sy+6* (1+abs(int($sy/6)));}
        else {
             $TT_val = $T_val;}

        # The directions for township and range are the same with those of
sampling site
        # Need more analysis here
        $TT_dir = $T_dir;
        $RR_dir = $R_dir;

        # then, get the section number from its xy coordinates
        if($sy==1 or $sy==3 or $sy==5) {
           $SS = $sy*6 - $sx + 1;}
        else {
           $SS = ($sy-1)*6 + $sx;}

        # now, get the MTRS
        if ( $TT_val <=9 ) {$TT_val = "0"."$TT_val";}
        if ( $RR_val <=9 ) {$RR_val = "0"."$RR_val";}
        if ( $SS <=9 ) {$SS = "0"."$SS";}

        #$new_MTRS[$DX+$i+1][$DY+$j+1] =
"M"."$TT_val"."$TT_dir"."$RR_val"."$RR_dir"."$SS";
        #print TEMP_OUT "$new_MTRS[$DX+$i+1][$DY+$j+1] ";

        $new_MTRS[$k] = "M"."$TT_val"."$TT_dir"."$RR_val"."$RR_dir"."$SS";
        $new_DIST[$k] = sqrt($i**2 + $j**2);
        if ( $new_DIST[$k]==0) {$new_DIST[$k] = .5;}
        print MTRS_OUT "$new_MTRS[$k] ";
        print DIST_OUT "$new_DIST[$k] ";
        $k++;


} # end of i loop

#print TEMP_OUT "\n";
#print  "\n";

} # end of j loop

close MTRS_OUT;
close DIST_OUT;
system ("cd ..");
return (\@new_MTRS, \@new_DIST);
} # end of sub
```

3. Output

ID	County	Station	Week	Conc	Appl(1)	Appl(2)	Appl(3)	Appl(4)	Appl(5)	Appl(6)	Appl(7)
1	Mont	SAL	1	1.64	0	6419	15253	36670	56451	56786	60608
1	Mont	SAL	2	2.36	0	0	2253	20970	32288	59092	63540
1	Mont	SAL	3	0.77	0	0	2053	26271	35052	75340	75340
1	Mont	SAL	4	0.5	0	3781	21193	34569	45571	80293	80293
1	Mont	SAL	5	0.7	0	9198	13581	40508	49757	64033	75364
1	Mont	SAL	6	3.01	0	0	14014	40599	46819	80605	90043
1	Mont	SAL	7	1.2	0	0	4652	14618	15787	35316	43891
1	Mont	SAL	8	0.14	0	0	0	0	0	3180	3180
2	Mont	OAS	1	0.38	0	0	0	0	0	0	0
2	Mont	OAS	2	0.44	0	0	0	0	0	0	0
2	Mont	OAS	3	0.17	0	0	0	0	0	0	0
2	Mont	OAS	4	0.4	0	0	0	0	0	0	1330
2	Mont	OAS	5	0.25	0	4730	4730	4730	4730	4730	7376
2	Mont	OAS	6	1.01	0	0	13720	13720	13720	13720	13720
2	Mont	OAS	7	0.39	0	0	0	0	0	0	0
2	Mont	OAS	8	0.08	0	0	0	0	0	0	0
3	Mont	CHU	1	0.73	0	0	2035	2035	2035	5427	5427
3	Mont	CHU	2	1.3	5394	5394	5896	12544	12544	12544	12544
3	Mont	CHU	3	0.34	0	0	0	10185	20330	29551	29551
3	Mont	CHU	4	0.4	4221	4221	17451	19126	31575	31575	31575
3	Mont	CHU	5	0.26	0	0	1968	2064	11317	17912	17912
3	Mont	CHU	6	1.61	0	0	10843	12424	17063	21372	33390
3	Mont	CHU	7	0.59	0	0	4688	4688	4688	4688	4688
3	Mont	CHU	8	0.11	0	0	0	0	0	0	0

……

Appendix III

A Perl program for linear regression model and its confidence intervals

1. Source file (linear.pl)

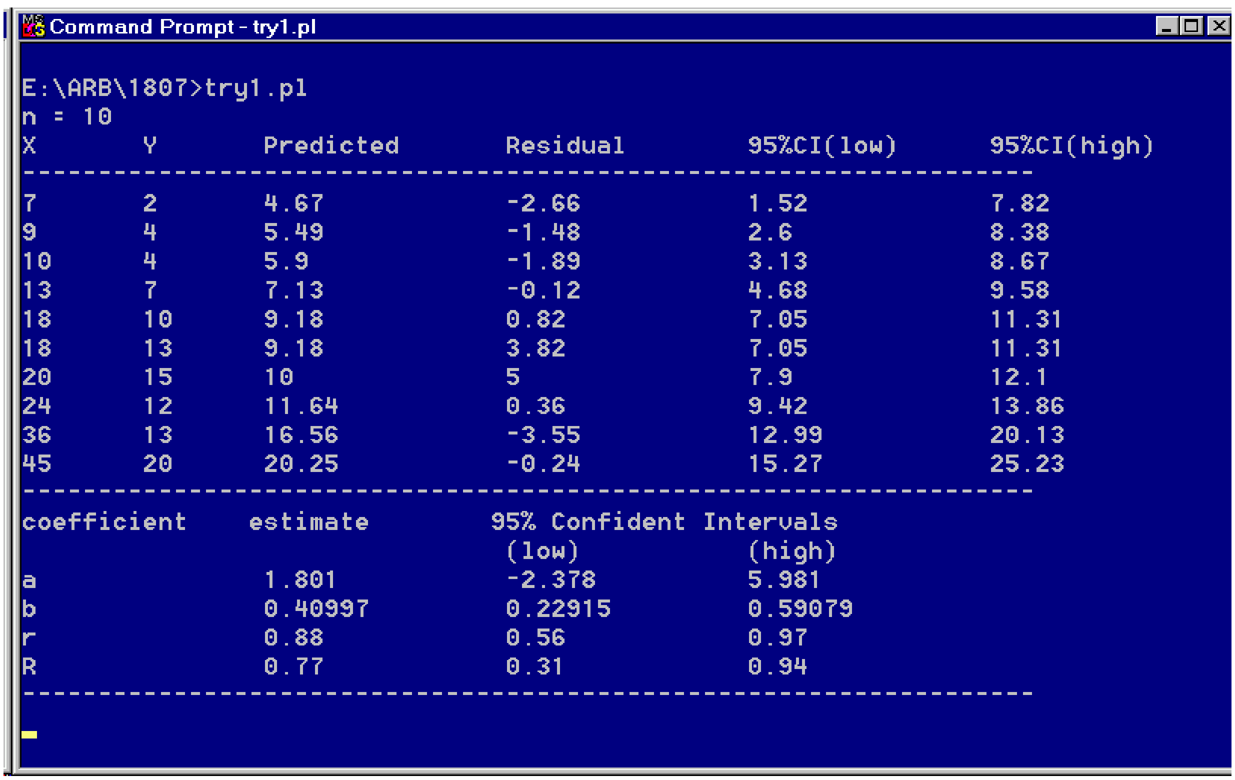
```
#!/usr/local/bin/perl
# Last change:  LI   30 Apr 2001   11:14 am
# linear.pl
# This module accepts (X,Y) pairs of data and does a linear regression: Y = a
+ bX.
# It calculates regression coefficients a, b and correlation coefficient r,
and their Confidence Intervals.
# It predicts Y values and their CIs for each X value.

# The formula and notation are from book "Statistical Methods for the Social
Sciences", P253-273

# testing data set from text book
#@X_array = (2,19,34,40,8,12,20,20,37,19,30,46);
#@Y_array = (48,21,14,11,41,37,22,31,19,42,15,18);
@X_array = (7,9,10,13,18,18,20,24,36,45);
@Y_array = (2,4,4,7,10,13,15,12,13,20);
 $Ref_Xarray = \@X_array;
 $Ref_Yarray = \@Y_array;
sub_Linear($Ref_Xarray, $Ref_Yarray);

# ------------------------------------------------------------ #
sub sub_Linear {
# usage: sub_linear($X_ref, $Y_ref)
my ($X_ref, $Y_ref) = @_;

# look-up table of t values for alfa = 0.050 and 0.025
@t_0050=(0.000, 6.314, 2.920, 2.353, 2.132, 2.015, 1.943, 1.895, 1.860, 1.833,
1.812,1.796, 1.782, 1.771, 1.761, 1.753, 1.746, 1.740, 1.734, 1.729, 1.725,
1.721, 1.717, 1.714, 1.711, 1.708, 1.706, 1.703, 1.701, 1.699, 1.645);

@t_0025=(0.000,12.706, 4.303, 3.182, 2.776, 2.571, 2.447, 2.365, 2.306, 2.262,
2.228,2.201, 2.179, 2.160, 2.145, 2.131, 2.120, 2.110, 2.101, 2.093, 2.086,
2.080, 2.074, 2.069, 2.064, 2.060, 2.056, 2.052, 2.048, 2.045, 1.960);


# read (X,Y) data pairs
@X = @{$X_ref};
@Y = @{$Y_ref};


# initializing variables
$n= @X;
$X_sum =0;
$Y_sum =0;
$XX_sum =0;
$XY_sum =0;

# calculate $a and $b
for ($i=0;$i<=$n-1;$i++)
```

```
{
    $X_sum += $X[$i];
    $Y_sum += $Y[$i];
    $XX_sum += $X[$i]*$X[$i];
    $XY_sum += $X[$i]*$Y[$i];
    $YY_sum += $Y[$i]*$Y[$i];
}

$X_ave = $X_sum/$n;
$Y_ave = $Y_sum/$n;

$b =  ($XY_sum - $X_sum * $Y_sum/$n) / ($XX_sum - $X_sum**2/$n);
$a =  $Y_ave - $b*$X_ave;


# SSE (sum of squared errors), EMS (error mean sqaure) and r, R
$SSE= 0;
$SSX = 0;
$SSY = 0;

for ($i=0;$i<=$n-1;$i++)
{
    $SSE += ($Y[$i]-($a+$b*$X[$i]))**2;
    $SSX += ($X[$i] - $X_ave)**2;
    $SSY += ($Y[$i] - $Y_ave)**2;
}

$Sigma = sqrt($SSE/($n-2));
$SigmaX = sqrt($SSX/($n-1));
$SigmaY = sqrt($SSY/($n-1));

$EMS = $SSE /$n;
$r = ($SigmaX/$SigmaY)*$b;


# confidence interval for regression coefficient b
$Sigma_b = $Sigma / sqrt($SSX);
#$t1 = 2.262;  # the t value is not a constant, should change with df=n-2 and
alfa value
               # in this case, n=11, df =9, alfa = 0.05 (for 95% CI)
               # t_0.025(9) = 2.262 in Page 528
$df = $n-2;
if ( $df >=30 ) { $df=30 ;}
$t = $t_0025[$df];
$b1 = $b - $t * $Sigma_b;
$b2 = $b + $t * $Sigma_b;


for ($i=0;$i<=$n-1;$i++)
{
    $Y0[$i] = $a+$b*$X[$i];
    $DY[$i] = $Y[$i] - $Y0[$i];
    $Y1[$i] = $Y0[$i] - $t*$Sigma*sqrt(1/$n + ($X[$i]-$X_ave)**2/$SSX);
    $Y2[$i] = $Y0[$i] + $t*$Sigma*sqrt(1/$n + ($X[$i]-$X_ave)**2/$SSX);
    $Y0[$i] = int($Y0[$i]*100+.5)/100;
    $DY[$i] = int($DY[$i]*100+.5)/100;
    $Y1[$i] = int($Y1[$i]*100+.5)/100;
    $Y2[$i] = int($Y2[$i]*100+.5)/100;
```

```
}

# confidence interval for regression coefficient a
# obtained by the CI of Y when X = 0 (letting $X[$i] = 0 in above equations)
    $a1 = $a + $b*0 - $t*$Sigma*sqrt(1/$n + (0-$X_ave)**2/$SSX);
    $a2 = $a + $b*0 + $t*$Sigma*sqrt(1/$n + (0-$X_ave)**2/$SSX);

# 95% confidence interval for correlation coefficient r
# see page 271-274 and Table E on page 533
$Tr = 1.151 *log((1+$r)/(1-$r))/log(10); # In Perl log(expr) returns natural
logarithm of expr;
$SigmaT = 1/sqrt($n-3);                  # logX = lnX/ln10
$Tr1 = $Tr - 1.96 * $SigmaT;
$Tr2 = $Tr + 1.96 * $SigmaT;

$r1 = (10**($Tr1/1.151)-1) / (10**($Tr1/1.151)+1);
$r2 = (10**($Tr2/1.151)-1) / (10**($Tr2/1.151)+1);

$R = $r**2;
$R1 = $r1**2;
$R2 = $r2**2;

if ( $r1<0 and $r2>0){
     if ($R1>$R2) {$R2 = $R1;}
     $R1=0;
   }
if ($r1<0 and $r2<0){
    $tmp = $R2;
    $R2 = $R1;
    $R1 = $tmp}


# need to convert Tr1 and Tr2 from table E on page 533 to r1 and r2

# formating for print
$a = int($a*1000+.5)/1000;
$a1 = int($a1*1000+.5)/1000;
$a2 = int($a2*1000+.5)/1000;

$b = int($b*100000+.5)/100000;
$b1 = int($b1*100000+.5)/100000;
$b2 = int($b2*100000+.5)/100000;

$Tr = int($Tr*100+.5)/100;
$Tr1 = int($Tr1*100+.5)/100;
$Tr2 = int($Tr2*100+.5)/100;

$r = int($r*100+.5)/100;
$r1 = int($r1*100+.5)/100;
$r2 = int($r2*100+.5)/100;
$R = int($R*100+.5)/100;
$R1 = int($R1*100+.5)/100;
$R2 = int($R2*100+.5)/100;


# print original data pairs and analysis results
print "n = $n\n";
```

```
print "X\tY\tPredicted\tResidual\t95%CI(low)\t95%CI(high)\n";
print "-------------------------------------------------------------------\n";
for ( $i=0;$i<=$n-1;$i++)
{
    print "$X[$i]\t$Y[$i]\t$Y0[$i]\t\t$DY[$i]\t\t$Y1[$i]\t\t$Y2[$i]\n";
}
print "-------------------------------------------------------------------\n";
print "coefficient    estimate        95% Confident Intervals\n";
print "                                (low)           (high)\n";
print "a\t\t$a\t\t$a1\t\t$a2\n";
print "b\t\t$b\t\t$b1\t\t$b2\n";
#print "Tr\t\t$Tr\t\t$Tr1\t\t$Tr2\n";
print "r\t\t$r\t\t$r1\t\t$r2\n";
print "R\t\t$R\t\t$R1\t\t$R2\n";
print "-------------------------------------------------------------------\n";
$e = <>;

print OUT  "$a\t\t$a1\t\t$a2\t \t$b\t\t$b1\t\t$b2\t \t$R\t\t$R1\t\t$R2\n";
#print "$X_sum\t$Y_sum\t$XX_sum\t$XY_sum\n$Sigma\n";

} # end of sub
```

2. Example



```
Command Prompt - try1.pl

E:\ARB\1807>try1.pl
n = 10
X       Y       Predicted       Residual        95%CI(low)      95%CI(high)
-------------------------------------------------------------------
7       2       4.67            -2.66           1.52            7.82
9       4       5.49            -1.48           2.6             8.38
10      4       5.9             -1.89           3.13            8.67
13      7       7.13            -0.12           4.68            9.58
18      10      9.18            0.82            7.05            11.31
18      13      9.18            3.82            7.05            11.31
20      15      10              5               7.9             12.1
24      12      11.64           0.36            9.42            13.86
36      13      16.56           -3.55           12.99           20.13
45      20      20.25           -0.24           15.27           25.23
-------------------------------------------------------------------
coefficient    estimate        95% Confident Intervals
                                (low)           (high)
a               1.801           -2.378          5.981
b               0.40997         0.22915         0.59079
r               0.88            0.56            0.97
R               0.77            0.31            0.94
-------------------------------------------------------------------
```

References

Agresti Finlay, 1986. Statistical Methods for the Social Sciences, 2nd Edition, Dellen Macmillan. P253-273.

Appendix IV

A FORTRAN program to test the significance level of area regressions using a randomization procedure

1. Important notes. TESTC.FOR takes the file area.dat as input. This file must be correctly formatted and contains the concentration and use data for 1x1,3x3,.... User inputs comment and specifies which column (1-8) to analyze, with col 1= 1x1, col2=3x3, etc. Output consists of several statistics, last column is computed F value for each randomization.

2. **TESTC.FOR** listing

```
     program testc
cccccccccccccccccccccccc
c 6/14/101
c test the rega subroutine
c read in the entire matrix of conc, and pur/area values
c
c analyze them all
c
c these values calculated by rega seem to be correct compared to
minitab analysis
c also, i ran 10 randomizations and loooked at results, the
randomization appears to
c be working ok and the analysis is correct
c
ccccccccccccccccccccccccccc
     implicit none
     real
xm(20,100),x(100),y(100),m,c,r2,rsq,sumx2,xbar,ybar,regssq
     real f
     integer n,i,j,k,l
     real rx,ry,temp
     integer i1,i2
     character*80 st

     open(unit=1,status='old',file='area.dat')
     do 10 i=1,11
      read(1,15)y(i),(xm(i,j),j=1,8)
15    format(f5.0,f8.0,7f9.0)
10   continue
     close(1)
     open(unit=2,status='new',file='testc.out')
     write(0,223)
223  format('enter comment ')
     read(0,224)st
224  format(a80)
```

```
      write(2,225)st
225   format(a80)

c pick one to analyze
      write(0,7979)
7979  format(1x,'enter column number 1-8 to analyze ')
      read(0,7980)k
7980  format(i1)

      do 25 i=1,11  !load in the use from selected column
25    x(i)=xm(i,k)

      n=11

       do 111 j=1,11     !write base data into file
        write(2,112)x(j),y(j)
112     format(1x,f10.1,f10.2)
111    continue


      write(2,11)
11    format(1x,'    m,         c,           r2,             rsq,'
     1'       ',
     1' sumx2,      xbar,    ybar',
     1'    regssq,        f')


c now warm up random number generator

      do 200 i=1,500
200   call random_number(rx)

c loop to generate permutations and calculations
      do 400 l=1,10000
      do 220 i=1,100  !100 times should randomize the array
       call random_number(rx)
       call random_number(ry)
        i1=1+int(11.*rx)
221     i2=1+int(11.*ry)
        if(i1.gt.11.or.i1.lt.1.or.i2.gt.11.or.i2.lt.1)then
           write(0,799)i1,i2
799        format(1x,'bad index i1 or i2: ',i3,i3)
           stop
         endif
         if(i1-i2.eq.0)then
           call random_number(ry)
           goto 221
         endif
```

```
c switch 2 values
       temp=x(i1)
       x(i1)=x(i2)
       x(i2)=temp
220  continue

c x array randomized, now perform regression

      call rega(X,Y,N,M,C,R2,RSQ,SUMX2,XBAR,YBAR,REGSSQ)
      f=9.*regssq/rsq
      write(2,100)m,c,r2,rsq,sumx2,xbar,ybar,regssq,f
100   format(1x,9g12.3)
c     do 111 j=1,11     !used this to look at random values and
check rslts
c      write(2,112)x(j),y(j)
c112   format(1x,f10.1,f10.2)
c111  continue

400  continue
     stop
     end

     SUBROUTINE REGA(X,Y,N,M,C,R2,RSQ,SUMX2,XBAR,YBAR,REGSSQ)
c$DEBUG:
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
C
C PERFORMS REGRESSION ON X,Y PAIRS IN X,Y OF SIZE N
C
C RETURNS SLOPE IN M, ADDITIVE CONSTANT IN C, AND RSQUARED
C VALUE IN R2
C
C MODIFIED 1/12/93 TO RETURN
C RESIDUAL SUM OF SQUARES (RSQ), CORRECTED SUM OF SQUARES
C OF X (SUMX2=SUM((X-XBAR)**2))
C MEAN X,  (XBAR) MEAN OF Y (YBAR)
C
CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
     IMPLICIT INTEGER(A-Z)
     REAL X(N),Y(N),M,C,R2
     REAL XBAR,YBAR,RSQ
     REAL SUMX,SUMY,SUMXY,SUMX2,SUMY2
     REAL TEMP,TEMP2
     REAL REGSSQ

c    write(0,5443)x(n),y(n),n
c5443     format(1x,'x(n),y(n),n',2f10.3,i4)

     SUMX=0.
```

```
      SUMY=0.
      SUMXY=0.
      SUMX2=0.
      SUMY2=0.
      DO 10 I=1,N
       SUMX=SUMX+X(I)
       SUMY=SUMY+Y(I)
       SUMXY=SUMXY+X(I)*Y(I)
       SUMX2=SUMX2+X(I)*X(I)
       SUMY2=SUMY2+Y(I)*Y(I)
10    CONTINUE

c     write(0,2525)sumx,sumy,sumxy,sumx2,sumy2
c2525      format(1x,'sumx,sumy,sumxy,sumx2,sumy2',/1x,5g15.6)

      TEMP=SUMXY-(SUMX*SUMY/FLOAT(N))
      M=TEMP/(SUMX2-(SUMX*SUMX/FLOAT(N)))
      C=(SUMY-M*SUMX)/FLOAT(N)
      TEMP2=SUMY2-(SUMY*SUMY/FLOAT(N))
      R2=M*TEMP/TEMP2
      XBAR=SUMX/FLOAT(N)
      YBAR=SUMY/FLOAT(N)
      SUMX2=SUMX2-(SUMX*SUMX)/FLOAT(N)
      REGSSQ=M*(SUMXY-SUMX*SUMY/FLOAT(N))
      RSQ=SUMY2-(SUMY*SUMY)/FLOAT(N)-REGSSQ
      RETURN
      END
```

3. **AREA.DAT**

0.19	0	0	0	0	0	45	45	45
2.16	0	955	9448	9448	9448	9448	15085	15085
0.09	0	0	0	0	0	0	32	77
0.79	0	0	0	955	9448	9448	9448	9448
0.10	0	0	0	0	0	0	0	32
0.67	0	1202	1202	5360	7883	12444	15384	16886
3.88	1114	6259	8590	23630	31427	40985	56066	62146
0.39	0	0	591	2306	2306	2306	2306	2803
1.29	0	0	4352	14848	26821	46393	59637	63250
7.73	2232	13633	26326	51372	66951	77251	80595	82312
2.61	169	6629	14111	24311	39339	55178	60308	66495

4. Results

Table APPENDIX IV. Results of randomization test for 8 area regressions.

area	f value	row	cumulative fraction
1x1	79.36	9992	0.9992
3x3	99.75	9998	0.9998
5x5	93.03	10000	1
7x7	157	10000	1
9x9	60.8	9996	0.9996
11x11	25.94	9990	0.999
13x13	20.47	9993	0.9993
15x15	17.99	9994	0.9994